

# Stock Price Prediction Using Machine Learning Techniques

---

Machine Learning to Explore Stock Price Prediction

**Mason**

**5/22/2021**

The world of stock trading is vastly complex. This project's aim is to build a machine learning model to reduce the decision-making effort when evaluating a stock's performance by analyzing historical data to predict future trends.

## Table of Contents

A. Project Overview.....	2
A1. Research Question or Organizational Need .....	2
A2. Context and Background .....	2
A3. Summary of Published Works.....	2
A4. Summary of Data Analytics Solution .....	4
A5. Benefits and Support of Decision-Making Process.....	4
B. Data Analytics Plan.....	5
B1. Goals, Objectives, And Deliverables .....	5
B2. Scope of Project .....	5
B3. Standard Methodology .....	5
B4. Timeline and Milestones.....	6
B5. Resources and Costs .....	7
B6. Criteria for Success .....	7
C. Design of Data Analytics Solution .....	8
C1. Hypothesis .....	8
C2. Analytical Method.....	8
C2a. Justification of Analytical Method .....	8
C3. Tools and Environments of Solution.....	8
C4. Methods and Metrics to Evaluate Statistical Significance .....	9
C4a. Justification of Methods and Metrics .....	9
C5. Practical Significance.....	9
C6. Visual Communication .....	10
D. Description of Datasets .....	10
D1. Source of Data .....	10
D2. Appropriateness of Dataset.....	10
D3. Data Collection Methods .....	10
D4. Data Quality .....	10
D5. Data Governance, Privacy and Security, Ethical, Legal, And Regulatory Compliance .....	11
D5a. Precautions .....	11
E. Sources .....	11

## A. Project Overview

### A1. Research Question or Organizational Need

This project will automate the process of data gathering and stock analysis. Analysis in this context is the application of machine learning techniques to predict future stock prices based on historical data.

### A2. Context and Background

Researching a company's performance and analyzing trends can be a time consuming and tedious task. By automatically downloading stock data and performing analysis using machine learning techniques, the time-consuming and overwhelming decision on whether to buy a particular stock can be greatly reduced.

### A3. Summary of Published Works

In Robert Kissell's book, *Algorithmic Trading Methods* (1), Algorithmic Trading is the process by which stock analysis, decision making, and trade execution are performed. In some markets, such as the foreign exchange market, it is estimated that 92% of trades are executed via AT rather than people (1). Many algorithms attempt to capitalize on the variability of a stock within a market day. Known as day trading algorithms, trading that uses the intra-day volatility of the stock market is estimated to comprise 70%-80% of all trades in financial markets.

The use of automated trading tools, machine learning, and artificial Intelligence is on the rise in now more than ever in the marketplace. An ever-growing need is placed on people with the knowledge to assemble, run, test, and analyze the tools used in the world digital trading.

The Institute of Electrical and Electronics Engineers (2), known as the IEEE, published a paper outlining how four different machine learning algorithms stacked up against one another when analyzing 463 stocks for the S&P 500. While it was difficult for historical data alone to predict future performance, it was noted that the addition of outside information, such as other indexes in foreign markets, yielded a larger increase in predictability.

Stock variability is not influenced by history alone. Foreign policy, investor psychology, and world events among other variables are all large influences and have a place in the prediction of market performance. While the techniques used by the IEEE used were advanced, it should be noted that a more holistic approach to the markets should be used when making stock decisions.

In an article published on Money Control, a popular finance news source, Kanika Awarrrwal referenced a quote from Warrant Buffet saying that “What's needed is a sound intellectual framework for making decisions and the ability to keep emotions from corroding that framework,” when speaking of trading strategies. She argues that “The best way to keep emotions out of investing is the use of technology. (3)” AI and Machine learning lend investors to remove the emotion from their decisions. By algorithmically analyzing the market, stronger financial decisions can be made and used to effectively leverage your financial position.

Humans are notoriously susceptible to emotional decision making. A person's moods, stress level, quality of sleep, and even hunger have to ability to influence decision making ability. By removing these barriers to quality decisions using machine learning, people can regain the ability to make sound decisions regardless of the underlying influences on their rational behavior.

#### A4. Summary of Data Analytics Solution

My objective is to create a standalone Python Application that will automate the process of ETL (Extract, Transform, Load) for ease of data manipulation. The application will take the data and perform regression analysis and other machine learning techniques to build a predicted stock price for a future date. The application will take ticker symbols such as MSFT (Microsoft) and download stock data spanning over a user-specified period. The application will also output graphics to aid in human-decision making.

#### A5. Benefits and Support of Decision-Making Process

My firm will benefit in the time. With the process of data-downloading being automated, users will not have to manually find and download data. Additionally, data transformation such as XLX (Microsoft Excel File) to CSV (Comma Separated Value) will be behind the scenes from the user's perspective.

The time previously spent manually analyzing historical stock data will be greatly simplified. A future stock price will be predicted allowing for decision makers at my Firm to quickly make critical decisions without the lag that accompanies "hand-jammed" analysis.

## B. Data Analytics Plan

### B1. Goals, Objectives, And Deliverables

The goal of this project is to create a Python application that will download stock data, perform ETL procedures, and utilize machine-learning techniques to predict a stock's future price. This goal can be broken down into the following objectives:

1. Download data from a stock of choice
  - The deliverable for this objective will be successfully downloaded stock data.
2. Transform the data into a format easy to interpret and manipulate with Python
  - The deliverable for this will be a Panda's Data Frame of stock data over a specified period.
3. Perform machine learning techniques to output a predicted value
  - The deliverable for this will be a predicted value.
4. Display predicted output in a user-friendly manner to aid in decision making
  - The deliverable for this will be applicable charts or graphs containing historical data and predicted values.

### B2. Scope of Project

The scope of this project will be to create a standalone Python application that takes in ticker symbols as its input. The output will be applicable charts or graphs as well as a predicted stock value for a future date.

### B3. Standard Methodology

Because this project will be created by one person and the aim is a simplified approach, this project will take on the Waterfall project methodology. This methodology is the best fit given this project's

circumstances. Project execution and implementation will adhere to the Waterfall Methodology's phases that will be discussed below.

Requirements: The requirements of this project will be a predicted stock value and graphics to aid in user interpretation.

Design: Project deliverables from section B1 will be mirrored for this phase. I will need to design a way to download stock data, render the data in a useful format for manipulation, perform analysis, and output the data in a user-friendly manner.

Implementation: During this phase, I will complete the deliverables for the goals listed in section B1.

Verification: I will use historical data to validate the output of my program. The output of my program must be of a reasonable value with respect to the historical data of a particular stock.

Maintenance: This phase will be on an as-needed basis and will be used to ensure that the application remains able to run given updates to Operating Systems, Python, or the Python libraries used.

#### B4. Timeline and Milestones

Milestone	Start Date	End Date	Duration (hours/days)
Establish project requirements	31 May 2021	31 May 2021	1 Day
Create module for Objective 1	31 May 2021	01 June 2021	2 Days
Create module for	01 June 2021	02 June 2021	2 Days

Objective 2			
Create module for Objective 3	02 June 2021	03 June 2021	2 Days
Create module for Objective 4	03 June 2021	04 June 2021	2 Days
Combine prior modules into one application	04 June 2021	05 June 2021	2 Days

## B5. Resources and Costs

Resources and/or Personnel	Cost
Obtaining Stock Data	N/A
Python 3.x Programming Environment	N/A
Python Modules	N/A
6 Workdays (Billed at \$75.00 Per Workday)	\$450.00

## B6. Criteria for Success

The criteria for success will be an application that is able to take a specified ticker symbol and output a projection for that stock's future price. The predicted value must fall within a 15% margin of error with respect to the true value. Additionally, this data must be graphically represented for ease of human



interpretation. If these conditions are met, then this project will be considered to have successfully met its requirements.

## **C. Design of Data Analytics Solution**

### **C1. Hypothesis**

My Machine Learning Model will forecast a stock's closing price in order to facilitate informed decisions on whether to buy or sell based on past performance.

### **C2. Analytical Method**

The Predictive Analytical Method is what will be used for this project. I will use data aggregation as the data analysis technique. Support Vector Machines (SVM) are supervised learning models. They excel at both data classification and regression. My application will use a SVM algorithm in order to predict a future price.

#### **C2a. Justification of Analytical Method**

This method was chosen because the goal of this project is to predict a future price based on historical data. Predictive analytics is best used when a description of the future is the objective. As such, it is fitting that a Support Vector Machine will be utilized for this project. SVM's are able to take in time-series data in order to forecast future values. This makes them suitable for my project.

### **C3. Tools and Environments of Solution**

For this project, a Python will be used to gather data, perform ETL procedures, provide analysis, and to output a predicted stock price. I will use common Python modules such as Numpy and Pandas to accomplish this. Additionally, I will use the SVM module from the Sci-Kit Learn library in order to make my machine learning model.

## C4. Methods and Metrics to Evaluate Statistical Significance

The metric to measure the performance of my application will be the predicted price. The predicted price will be a number that is expected to accurately predict a future value. I wish to test whether the observed predicted value is within a 15% range of the true value for a particular day. For example, if the true price is \$100 and the predicted price is \$105, the predicted price would be 5% greater than the true price. In this case, the value falls within a 15% range and would be considered successful.

### C4a. Justification of Methods and Metrics

By using the above metric to measure performance, I will be able to prove my models' accuracy within a chosen tolerable margin.

## C5. Practical Significance

The practical significance of this tool will allow my firm to automate tasks that would have otherwise been manually performed. For example, if downloading stock data takes 10 minutes to complete when manually performed and it takes 8 minutes to complete when performed with my tool, that is a time savings of 2 minutes per action.

Performing analysis on a stock is a bit more difficult to estimate but we can make some assumptions to simplify this. Suppose that manually performing analysis on a stock based on historical data would take 20 minutes when manually performed. This can be automated with machine learning techniques.

However, machine learning efficiency can be increased by applying fit models. Only the first iteration, or first series of iterations will take the most time to complete. For other iterations, the model can be fit to optimize future analyzations. Future analyzations might, for example, take only 3 minutes. Time savings in this case would be approximately 19 minutes per analyzation.

## C6. Visual Communication

I will graphically display the difference between a stocks true value against my models' predicted value. I will take the difference, if any, between the values and fit them to a normal distribution curve. I will then find the standard deviation to produce a Z score. This Z score will then be used to either reject or accept the null hypothesis.

## D. Description of Datasets

### D1. Source of Data

The source of my data will be publicly available stock data. The dataset will be pulled from the Yahoo! Finance API and Google Finance API as appropriate.

### D2. Appropriateness of Dataset

To meet the goals and objectives of this project, I will need stock data. This data is appropriate because to predict a future value, an accurate sample of historical values is needed.

### D3. Data Collection Methods

To collect the data necessary for this data set, I will use publicly available stock information spanning a variety of stocks. I will SPY stock data as a proxy for the S&P 500. Furthermore, I will pull samples of Blue-Chip Stocks from within the S&P 500. Examples include large, publicly traded companies such as Microsoft (MSFT), Google (GOOG), Tesla (TSLA), Apple (AAPL), and Facebook (FB).

### D4. Data Quality

Publicly available stock datasets are largely complete and free from error. Any null, Nan, or zero-values will be removed from this dataset to remove the presence of outliers.

## D5. Data Governance, Privacy and Security, Ethical, Legal, And Regulatory

### Compliance

Because of the public availability of stock data through various online resources, there exists no concern over the privacy, security, ethicality, legality, or regulatory compliance my dataset.

#### D5a. Precautions

The world of stocks and financial markets is wildly complex. This project attempts to reduce one aspect of that complexity by analyzing historical data. Judgements should not be made on this project's output alone. Any output from this project should not be construed as financial advice nor does it advocate for the investment of any securities. Never invest more than you are comfortable with losing.

### E. Sources

1. Kissell, R. (2020, September 4). Algorithmic Trading Methods. Algorithmic Trading Methods - 2nd Edition. <https://www.elsevier.com/books/algorithmic-trading-methods/kissell/978-0-12-815630-8>
2. Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks. IEEE Xplore. (n.d.). <https://ieeexplore.ieee.org/abstract/document/8258518>.
3. Awarwal, K. (2021, May 22). AI, machine learning help investors gain 'edge' while investing. Money Control. <https://www.moneycontrol.com/news/business/markets/ai-machine-learning-help-investors-gain-edge-while-investing-6922331.html>